

# L03 – Introduction to Graphical Models

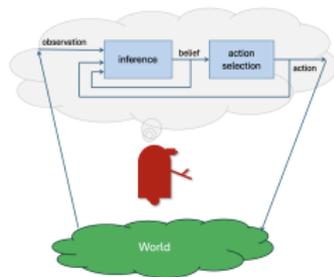
AIMA4e, 13.1–2

# What you should know after this lecture

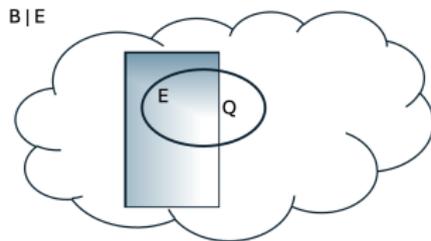
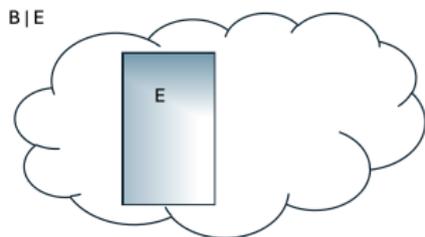
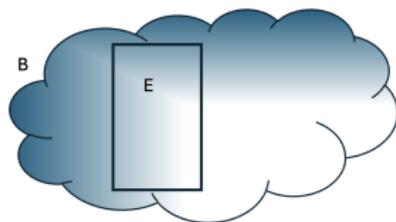
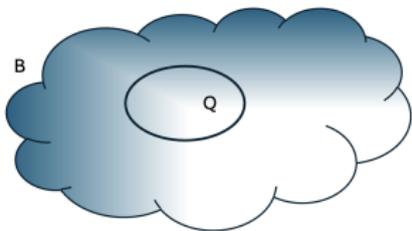
- Framing of probabilistic inference problem
- How to model a distribution of variables as a factored distribution
- How to represent a factored distribution as a graphical model
- How (and why) to multiply and marginalise out random variables

# Probabilistic belief representation

- Belief is a probability distribution :  
 $B \in \mathcal{P}(\mathcal{S})$   
(an element of the set of all distributions over  $\mathcal{S}$ )
- Important questions:
  - What is  $p_B(\text{event})$ ?
  - What is the most likely state  $\text{argmax}_s p_B(s)$ ?
  - How should we update  $B$  given an observation?



# Belief, query, conditioning



# Probabilistic inference

Given  $B$  and  $Q$  and possible  $E$ , compute

$$P_{B|E}(Q) = P_B(Q, E) / P_B(E)$$

Stupidest possible algorithm:

- Enumerate  $s \in \mathcal{S}$
- accumulate  $p_B(s)$  if  $s \in Q \cap E$
- accumulate  $p_B(s)$  if  $s \in E$
- Return ratio

Our goal: do this without enumerating  $\mathcal{S}$

Idea: use factored representation of  $B$ ,  $Q$ , and  $E$  to make this efficient!

## Factored representation of B

- Random variables  $V_1, \dots, V_n$
- Each  $V_i$  has discrete domain of possible values  $\Omega_{V_i}$
- Sample space is product  $\mathcal{S} = \Omega_{V_1} \times \dots \times \Omega_{V_n}$
- Sample  $s \in \mathcal{S}$  is  $(v_1, \dots, v_n)$  where  $v_i \in \Omega_{V_i}$
- B is the joint distribution on  $V_1, \dots, V_n$
- Can use a table  $\alpha$  to represent B
- Use Boolean expressions over atoms  $V = v$  to represent Q and E

## Factored representation: example

- Random variables  $A, B, C$
- Domains  $\Omega = \{0, 1\}$

	a	b	c	$p((a, b, c))$
$\alpha =$	0	0	0	0.10
	0	0	1	0.20
	0	1	0	0.05
	0	1	1	0.05
	1	0	0	0.30
	1	0	1	0.05
	1	1	0	0.15
	1	1	1	0.10

What is  $P_{\alpha}(A = 1 \mid B = 0 \text{ or } C = 0)$

# Bayes Nets: Compact factored representation of $p$

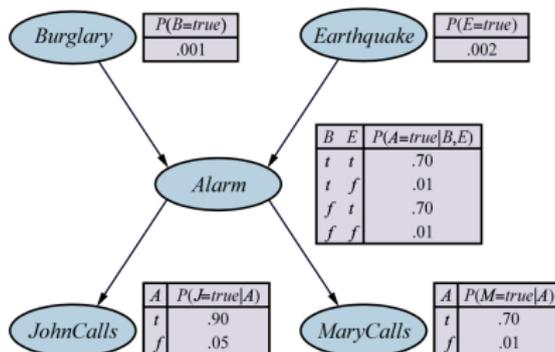
Define a Bayesian network  $\alpha$  :

- Random variables  $V_1, \dots, V_n$
- Each  $V_i$  has discrete domain of possible values  $\Omega_{V_i}$
- Directed acyclic graph  $G$  defined on nodes  $V_i$
- Parents  $\text{pa}_G(V_i)$  : set of nodes  $V_j$  with edges  $(V_j, V_i) \in G$
- For each  $V_i$ , a conditional probability table (CPT), specifying  $P(V_i \mid \text{parents}_G(V_i))$ 
  - For every assignment  $\bar{v}$  to variables in  $\text{pa}_G(V_i)$
  - and every value  $v \in \Omega_{V_i}$
  - specify  $P(V_i = v \mid \text{pa}_G(V_i) = \bar{v})$

Then for an assignment  $s = (v_1, \dots, v_n)$

$$p_\alpha(s) = \prod_i P(V_i = v_i \mid \text{pa}_G(V_i) = s[\text{pa}_G(V_i)])$$

# Classic example



**Figure 13.2** A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters  $B$ ,  $E$ ,  $A$ ,  $J$ , and  $M$  stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

- Non-monotonicity of probability
  - What's  $P_{\alpha}(B = 1)$ ?
  - What's  $P_{\alpha}(B = 1 \mid M = 1)$ ?
  - What's  $P_{\alpha}(B = 1 \mid M = 1, E = 1)$ ?
  - How many params to specify the whole joint as a table?

# Explaining away

Consider the network

Battery  $\rightarrow$  Gauge  $\leftarrow$  FuelTank

Here are some CPTs:

$$\Pr(B = 1) = 0.9$$

$$\Pr(F = 1) = 0.9$$

$$\Pr(G = 1 \mid B = 1, F = 1) = 0.8$$

$$\Pr(G = 1 \mid B = 1, F = 0) = 0.2$$

$$\Pr(G = 1 \mid B = 0, F = 1) = 0.2$$

$$\Pr(G = 1 \mid B = 0, F = 0) = 0.1$$

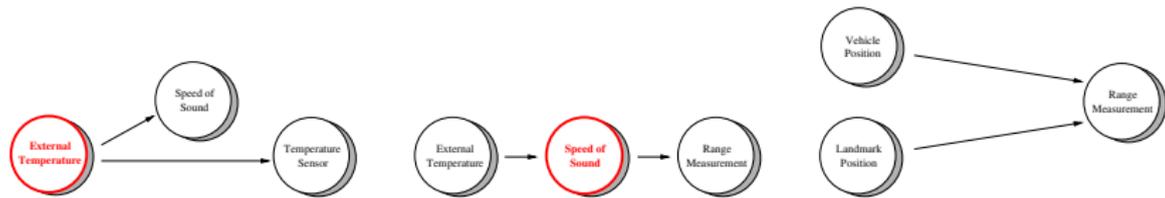
- What is the prior that the tank is empty?  $\Pr(F = 0) = 0.1$
- What if we observe the fuel gauge and find that it reads empty?  $\Pr(F = 0 \mid G = 0) \approx 0.257$
- Now, what if we find the battery is dead?  
 $\Pr(F = 0 \mid G = 0, B = 0) \approx 0.111$  The probability that the tank is empty has decreased! Finding that the battery is flat explains away the empty fuel tank reading.

# Independence relations

Are we getting something for nothing?

- Independence of random variables: If  $P(A = a, B = b) = P(A = a)P(B = b)$  for all  $a \in \Omega_a, b \in \Omega_b$ , we say that  $A$  and  $B$  are independent:  $A \perp\!\!\!\perp B$ .
- Conditional independence: If  $P(A = a, B = b \mid C = c, D = d) = P(A = a \mid C = c, D = d)P(B = b \mid C = c, D = d)$  for all  $a \in \Omega_A, b \in \Omega_B, c \in \Omega_C, d \in \Omega_d$ , we say that  $A$  and  $B$  are conditionally independent given  $C$  and  $D$ ,  $A \perp\!\!\!\perp B \mid C, D$ .
- Bayes nets get their compactness from independence assumptions encoded in the graph.

# Graph structure encodes independence relations



- Case 1:  $P(B|A), P(C|A)$  “outgoing” connection
  - $B \not\perp C$ , but  $B \perp C | A$
- Case 2:  $P(B|A), P(C|B)$  “flow” connection
  - $C \not\perp A$ , but  $C \perp A | B$
- Case 3:  $P(C|A, B)$  “incoming” connection
  - $A \perp B$ , but  $A \not\perp B | C$

In general  $V_i \perp V_j | E_1, \dots, E_K$  if there are no paths from  $V_i$  to  $V_j$  through outgoing or flow connections that are not blocked by  $E$  or through an incoming connection that is enabled by  $E$ . More about this when we get to factor graphs and Markov blankets.

## Simple inference algorithm

Given a BN, we have a conceptually (but not computationally) simple way to compute the joint

$$p_{\alpha}(s) = \prod_i P(V_i = v_i \mid \text{pa}_G(V_i) = s[\text{pa}_G(V_i)])$$

We can think of this as multiplying the CPTS in the Bayes net. Informally:

MULTIPLY( $D_1, D_2$ )

- 1  $\pi$  = table indexed by  $\Omega_{\text{vars}(D_1) \cup \text{vars}(D_2)}$
- 2 **for**  $\bar{v}$  in  $\pi$
- 3      $\pi(\bar{v}) = \text{lookup}(\bar{v}, D_1) \cdot \text{lookup}(\bar{v}, D_2)$
- 4 **return**  $\pi$

## Multiplication example

Given CPTs,  $D_1 = P(X_2|X_1)$  and  $D_2 = P(X_3|X_1)$ , defined over different variable sets:

$$D_1 =$$

$X_1$	$X_2$	P
T	T	0.1
T	F	0.9
F	T	0.9
F	F	0.1

$$D_2 =$$

$X_1$	$X_3$	P
T	T	0.9
T	F	0.1
F	T	0.1
F	F	0.9

$$\text{MULTIPLY}(D_1, D_2) =$$

$X_1$	$X_2$	$X_3$	P
T	T	T	$0.1 \times 0.9 = 0.09$
T	T	F	$0.1 \times 0.1 = 0.01$
T	F	T	$0.9 \times 0.9 = 0.81$
T	F	F	$0.9 \times 0.1 = 0.09$
F	T	T	$0.9 \times 0.1 = 0.09$
F	T	F	$0.9 \times 0.9 = 0.81$
F	F	T	$0.1 \times 0.1 = 0.01$
F	F	F	$0.1 \times 0.9 = 0.09$

What is the meaning of this multiplication?

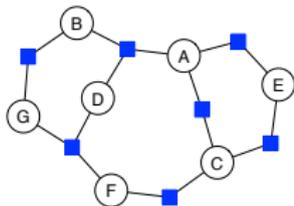
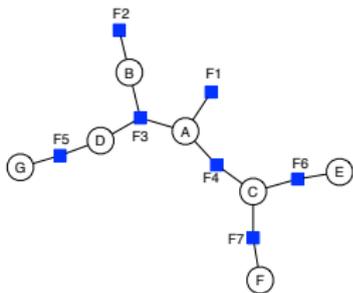
$P(X_2|X_1) \times P(X_3|X_1) = P(X_2, X_3|X_1)$ .

Practice on explaining-away example

# Undirected models

- Directed models (Bayes nets) are good for many problems, particularly when there is a causal interpretation of the arrows. (Though causality is not necessary)
- Relationship between pixels in an image or adjacent plots of property is not independent but there's no sensible way to assign a direction.
- Can make graphical models with nodes and undirected arcs: Markov random fields
- We will skip that step and go straight to a formalism called factor graphs that can represent both directed and undirected models.
- A generalization of factor graphs for CSPs!!

# Factor graphs



Undirected bipartite graph: factors only connect to variables

- Round nodes are random variables  $V$
- Square nodes are factors  $\phi$ : tables specifying, for each tuple of value of the connected variables, a non-negative number
- Represent a probability distribution (e.g. left graph above)

$$P((a, b, c, d, e, f)) = \frac{1}{Z} \phi_1(a) \phi_2(b) \phi_3(a, b, d) \phi_4(a, c) \phi_5(d, g) \phi_6(c, e) \phi_7(c, f)$$

where  $Z$  is a normalizer

$$Z = \sum_{a,b,c,d,e,f} \phi_1(a) \phi_2(b) \phi_3(a, b, d) \phi_4(a, c) \phi_5(d, g) \phi_6(c, e) \phi_7(c, f)$$

# Bayes nets to factor graphs

- Variable nodes are the same
- Add one factor for each CPT
- Connect it to the “output” node and all parents
- Note that, for this construction  $Z = 1$  (no need to normalize!)

Prove this to yourself by recalling the probability distribution represented by a Bayes net.

# Independence relations in factor graphs

- The Markov blanket of a node  $V$  consists of all nodes that are connected to any factor connected to  $V$ .
- The Markov blanket of  $A$  in our example is  $\{B, D, C\}$
- A node  $V$  is not, in general, independent of any node in its MB
- A node  $V$  is conditionally independent of the rest of the graph, conditioned on  $mb(V)$
- There are some sets of independence relations that are describable by a Bayes net but not describable by a factor graph (and vice versa)

# Inference in factor graphs

Some inference problems:

- Joint distribution: In a factor graph, use table multiplication to compute a big table

$$\frac{1}{Z} \prod_k \phi_k$$

where  $Z$  is the sum of all table entries

- Marginal distribution:  $P(Y)$  where  $Y \subset \mathcal{V}$
- Conditional probability:  $P(Y \mid E = e)$ , where  $Y \subset \mathcal{V}$ ,  $E \subset \mathcal{V}$ , and  $Y \cap E = \emptyset$ ; and  $e$  is the observed values of the variables in  $E$ . Note that it is not necessary that  $Y \cup E = \mathcal{V}$ .
- Most probable assignment (MAP):

$$\operatorname{argmax}_y P(Y = y \mid E = e) .$$

Note that the MAP of a set of variables is not necessarily the set of MAPs of the individual variables.

## PGMs and CSPs

- In both PGMs and CSPs, the nodes represent variables, with finite domains.
- The factors are tables of values, one for each assignment of the variables they are connected to.
- In CSPs, the values have to be 0 and 1.
- In PGMs, the values can be any non-negative number.
- The factor graph of a CSP defines a **set** of assignments.
- The factor graph of a PGM defines a **distribution** over assignments.

# Formulation

How can we model these problems?

- A collection of diseases and symptoms
- Whether a group of college students enjoy the music of Adam Ant
- A collection of attributes of a driver and their car and their likelihood of having a car crash

Two ways to think about evidence:

- Values of variables
- Additional factors

## Next time

- We would like to avoid computing the whole joint distribution!!
- Algorithms whose complexity depends on the complexity of the network (rather than the product of the domains of all the variables)